View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Calhoun, Institutional Archive of the Naval Postgraduate School

Calhoun

Institutional Archive of the Naval Postgraduate School

Calhoun: The NPS Institutional Archive

Faculty and Researcher Publications

Faculty and Researcher Publications

2013-02

Pilot Performance: Assessing How Scan Patterns & Navigational Assessments Vary by Flight Expertise

Yang, Ji Hyun

DUDLEY KNOX LIBRARY

Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

Pilot Performance: Assessing How Scan Patterns & Navigational Assessments Vary by Flight Expertise

Ji Hyun Yang, Quinn Kennedy, Joseph Sullivan, and Ronald D. Fricker, Jr.

Yang JH, Kennedy Q, Sullivan J, Fricker Jr RD. *Pilot performance: assessing how scan patterns and navigational assessments vary by flight expertise.* **Aviat Space Environ Med 2013; 84:1–9.**

Introduction: Helicopter overland navigation is a cognitively complex task that requires continuous monitoring of system and environmental parameters and many hours of training to master. This study investigated the effect of expertise on pilots' gaze measurements, navigation accuracy, and subjective assessment of their navigation accuracy in overland navigation on easy and difficult routes. ***Methods:*** A simulated overland task was completed by 12 military officers who ranged in flight experience as measured by total flight hours (TFH). They first studied a map of a route that included both easy and difficult route sections, and then had to 'fly' this simulated route in a fixed-base helicopter simulator. They also completed pre-task estimations and post-task assessments of the navigational difficulty of the transit to each waypoint in the route. Their scan pattern was tracked via eye tracking systems, which captured both the subject's out-the-window (OTW) and topographical map scan data. ***Results:*** TFH was not associated with navigation accuracy or root mean square (RMS) error for any route section. For the easy routes, experts spent less time scanning out the window ($\rho = -0.61$) and had shorter OTW dwell ($\rho = -0.66$). For the difficult routes, experts appeared to slow down their scan by spending as much time scanning out the window as the novices while also having fewer Map fixations ($\rho = -0.65$) and shorter OTW dwell ($\rho = -0.69$). However, TFH was not significantly correlated with more accurate estimates of route difficulty. ***Discussion:*** This study found that TFH did not predict navigation accuracy or subjective assessment, but was correlated with some gaze parameters.

Keywords: expertise, scan strategy, cognition, subjective assessment.

A COMMON GOAL IN training is to teach novices to behave and think like experts so that they can more quickly attain satisfactory levels of performance and decision-making skills (10). In aviation, performance is generally assessed by level of flight control, typically defined by root mean square (RMS) error of flight trajectory, accuracy of flight decisions, and depth of understanding of the issues surrounding the decision. Expert pilots, defined by total flight hours or FAA ratings, consistently perform these tasks better than less experienced pilots (1,9,12). Helicopter overland navigation is a particularly challenging aviation task for trainees and instructors as it entails additional cognitively demanding tasks above and beyond flight control. Furthermore, RMS error of flight trajectory does not predict expertise levels in helicopter overland navigation (15) as it does in other aviation tasks. This is because helicopter pilots are trained to adapt their between-waypoints navigation solution based on current observation. For example, pilots may elect to deviate from a straight-line connection between waypoints to take advantage of a guiding feature that was not readily apparent in preflight planning (15). Thus, in training helicopter pilots, a different measure of expertise beyond RMS error is needed.

Another limitation of using RMS error as a measure of flight expertise is that it does not provide information regarding experts' underlying cognitive strategies while flying or how these strategies may change with accrued experience. Currently, little is known about the learning process underlying improvements in flight control and navigation. For example, do experts simply demonstrate more precise control or do they do things in a qualitatively different way, by perhaps sampling different sources of information (1,7)? In order to better explain why pilots' overland navigation accuracy differ by expertise level and to find cues for assessing their cognitive states, we suggest observing human behaviors (e.g., where they look) which influence their performance (e.g., how they navigate). Even for one of the most common causes of mishaps, the breakdown in cockpit scan, developing a good scan strategy has not been given high priority during training (2). Standardized methods and scan patterns can be described to students at a high level and in general terms; however, actually assessing the appropriateness of a student's scan in relation to the in-situ training environment and their performance is not well supported. There is little support for instructors to provide carefully tailored feedback specific to a pilot and the immediate training environment. Thus, the purpose of this study was to attempt to understand underlying cognitive strategies used by experts that aid in superior performance.

The goal of our research is to improve training in challenging aviation tasks by providing instructors with real-time information regarding what the trainee is thinking. A large body of research demonstrates that eye scan parameters successfully predict different cognitive states, and we have found it to be able to detect an underlying cognitive strategy specific to overland navigation. Although navigation performance provides a

From the Naval Postgraduate School, Monterey, CA.

This manuscript was received for review in March 2012. It was accepted for publication in August 2012.

Address correspondence and reprint requests to: Ji Hyun Yang, Ph.D., 700 Dyer Rd., Bldg. 245, Rm. 265, Naval Postgraduate School, Monterey, CA 93943; jyan1@nps.edu.

Reprint & Copyright © by the Aerospace Medical Association, Alexandria, VA.

DOI: 10.3357/ASEM.3372.2013

marker of who completed the task well, it does not provide insight into the strategies that the person used to complete the task successfully.

Among psychophysiological measures for human cognitive states in real time, eye movements are relatively easy to collect in actual operational environments, and recent eye-tracking technology provides nonintrusive devices to collect ocular data (5). Sullivan et al. (15) gives descriptions of eye-tracking studies in several domains, including aviation (1,8), ground transportation (13), different cognitive states (11), and visual processing load (16). Thus, by knowing expert pilots' scan patterns for different aviation tasks and decisions, training novice pilots can be improved by 1) teaching them how to scan the environment more effectively; and 2) detecting experts' underlying cognitive strategies based on their scan patterns; these strategies can then be taught to novices.

Previous aviation studies that used eye-tracking did not investigate expertise and visual scan differences in helicopter overland navigation tasks, which are considered to be more cognitively demanding and continuously complex than fixed wing aircraft operating tasks. The distinguishing characteristic that makes it cognitively demanding is height above terrain. Given the altitude of helicopters above ground, pilots rely much more on terrain relief than their fixed-wing counterparts. From the higher altitude at which fixed-wing pilots fly, the visible terrain more closely matches the map representation, whereas helicopter pilots rely more on terrain relief at lower altitudes. Also, from higher altitudes more features are in view for a longer time. Recently, Sullivan et al. (15) demonstrated that when pilots were on track during an overland navigation task, flight expertise predicted gaze parameters and scan management skills, but did not predict flight performance measures, such as RMS error. However, it is unknown whether this pattern of results also occurs when pilots are faced with more difficult navigation routes in which they are more likely to be off track. It also is unknown how well experts' estimates of route difficulty match their actual performance. If experts know ahead of time which sections of the route are difficult to navigate, they may alter their visual scan strategies accordingly during these sections. From a training perspective, understanding expertise differences in the link between navigation accuracy, pilots' subjective assessment of how they are doing, and visual scan patterns would greatly enhance current training procedures. We thus focused on improving our understanding of cognitive processing associated with helicopter overland navigation by analyzing gaze measurements, navigation accuracy, subjective estimation and assessment, route difficulties, and expertise level of pilots.

In this study, we designed overland navigation tasks in a flight simulator integrated with eye-tracking systems and performed human-in-the-loop experiments with pilots who ranged in total flight hours. The simulated navigation tasks entailed 'flying' to 12 waypoints depicted on a map. In our previous work (15), we examined only waypoints 2–5, in which all pilots were on track. In this study, we extend upon these results by also examining expertise differences in actual and self-reported performance for waypoints 5–7, which were rated as much more challenging than waypoints 2–5. The Results section focuses on these two route sections; notable points from other waypoints data are described in the Discussion for an organized reporting.

Pre-experiment, we made the following alternative hypotheses for helicopter overland navigation tasks regarding route difficulty and expertise represented by total flight hours (TFH):

1. Correlation between TFH and navigation accuracy vs. TFH and RMS error: on both easy and difficult route sections,
   a. TFH will be positively correlated with navigation accuracy; and
   b. TFH will be correlated with RMS error.
2. Correlation between TFH and gaze parameters/scan patterns: on both easy and difficult route sections,
   a. Higher TFH will be associated with shorter out-the-window (OTW) dwell times;
   b. Higher TFH will be associated with shorter Map dwell times;
   c. Higher TFH will be associated with greater frequency of OTW fixations;
   d. Higher TFH will be associated with greater frequency of Map fixations;
   e. Higher TFH will be associated with a greater number of view changes between OTW and Map; and
   f. Higher TFH will be associated with less OTW scan duration.
3. TFH and subjective assessment:
   a. TFH will have a stronger correlation with the pre-survey than with actual navigation accuracy; and
   b. TFH will have a stronger correlation with the post-survey than with actual navigation accuracy.

As part of this project we developed a visualization tool, the Flight and Eye Scan visualization Tool, designed to provide a representation of spatial and temporal correspondence among features scanned in OTW (3D) and Map (2D) views in relation to the actual aircraft location (15).

METHODS

Subjects

There were 12 male military personnel, 29 to 40 yr of age, who participated in the study. The minimum skill requirement for the study was completion of at least one overland navigation class. Among the 12 subjects, 3 subjects were helicopter flight instructors and 2 subjects had other navigation-related instructing experience but no flight experience. Expertise was defined in terms of TFH, where higher TFH values were used as a proxy measure indicating increased pilot expertise. TFH varied from 0 to 3100 h (avg. = 1488 h, SD = 1104 h) and Q1
overland flight hours varied from 0 to 2500 h (avg. = 612 h, SD = 853 h). Seven subjects were from the U.S. Navy, two from the U.S. Marine Corps, one from the U.S. Air Force, one from the U.S. Army, and one was unknown. No special neurological, visual acuity, or spatial ability tests were performed. The study was approved by the Naval Postgraduate School Institutional Review Board. Subjects were recruited via an e-mail advertisement to Naval Postgraduate School e-mail account holders. All

the subjects gave written informed consent to participate, with the right to withdraw at any time.

Equipment

The basic experimental apparatus included the flight simulator X-Plane 8.6, a 46″ wide screen to present the OTW view, a 40″ wide display for the map and instrument display, two stereo cameras and associated faceLAB 4.6 software for collecting eye data, and a cockpit-style seat with side-mounted joystick. Data from the flight simulator were sent to an image generator, which provided an OTW and a map view combining an OpenSceneGraph terrain model of Twentynine Palms, CA. The helicopter was designed to be on an automated terrain-following mode at fixed 150′ above ground level flying at 60 kn. However, the pilot was able to control the heading of the aircraft using the lateral control of the joystick. The joystick pitch control (up/down) was programmed to change the up/down view of the OTW, not the actual pitch angle of the aircraft. The display presented a 1:50,000 topographical land map typically used for flight planning and execution. The map was fixed in position about the pair-wise mean of the waypoints, whereas the orientation of the map was synchronized to the aircraft's heading to maintain a track-up orientation. The bottom portion of the screen contained instruments to support navigation: the left-most instrument display was a compass typical of legacy Navy H-60 (SH/HH-60F/H) displays. To the right of the compass display were typical barometric and radar altimeters. The right-most portion of the instrument cluster contained a digital-style elapsed time clock. We had two separate faceLAB systems (two sets of stereo cameras with 12.5 mm lenses, three infrared strobe lights) for tracking eye gaze for OTW and map displays.

Procedures

The navigation task was to fly over 12 waypoints (indicated as black circles on **Fig. 1**) after studying the area using Falcon View flight planning software, a system widely employed by diverse communities within the U.S. Department of Defense. The first waypoint is located slightly south of the map so it is not shown in the figure. Each waypoint pair has a "doghouse" that indicates (from top to bottom): the next waypoint number, the recommended heading to reach that waypoint from the previous one, the distance between waypoints, and the amount of time it takes to traverse the distance assuming a speed of about 60 kn. Pilots were free to deviate as long as they remained oriented. Pilots were considered on-track when they stayed within 0.5 km from the waypoint and off-track when they deviated more than 0.5 km.

Waypoints were set very close together and the terrain tended to be ambiguous, so subjects needed to make course corrections based on visual cues from both the OTW and map screens (their goal being to bring their perceived location closer to their actual location). The task was purposely designed so that some legs would be more challenging than others. The difficulty of each leg was assessed by a subject matter expert (SME) who designed the whole route. The SME determined that the legs from waypoints 2–4 were easy, whereas the legs between waypoints 5–7 were difficult. We refer to waypoints 2–4 as the easy route section and waypoints 5–7 as the difficult route section. The two thick lines indicate the easy route section (waypoints 2–4) and the difficult route section (waypoints 5–7).

A pre-task questionnaire asked subjects to indicate the level of navigation difficulty for each of the 11 legs on a scale from 1 to 5, in which 1 = completely trivial, 2 = somewhat difficult, 3 = moderately difficult, 4 = very difficult, and 5 = not at all possible. The scale was represented as a straight horizontal line and participants were told to draw a vertical line to indicate their level of perceived difficulty. The numbers 1–5 were evenly dispersed above the line. Thus, perceived level of difficulty was calculated by measuring the marker distance from the left-most point of the scale (line) divided by the total length of the line. Multiplied by 100, self-reported perceived difficulty was thus quantified on a percentage scale.

A post-task questionnaire asked subjects to indicate the level of navigation difficulty that they experienced for each of the 11 legs on the same scale used for the pre-task questionnaire. A demographic survey had questions regarding subjects' age, gender, branch of military service, total flight hours, overland navigation hours, days since last flight, instructor experience, and years of aviation experience. After a brief introduction, subjects were asked to read and sign an informed consent form.

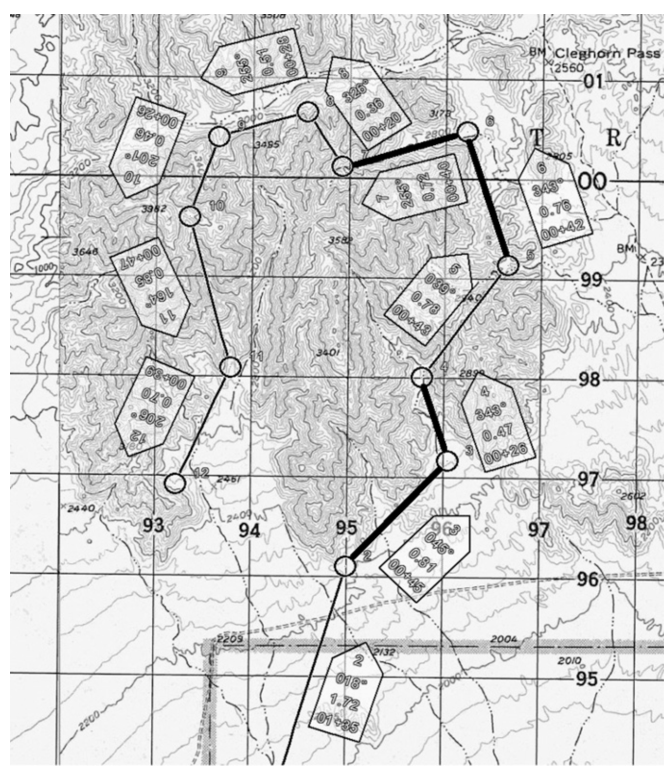

Fig. 1. Flight route showing 2nd to 12th waypoints with corresponding dog houses: waypoints 2-4 and 5-7 are shown in thick lines (15).

They then completed the demographic survey. The next step was a calibration of faceLAB stereo cameras to verify that the visual scan data was usable (error less than 3°) before subjects started the navigation tasks. Subjects were asked to sit in the simulator chair, where eye-tracking cameras had been mounted in between the chair and the simulator screen. Once the calibration was done, the simulated flight environment was explained to the subjects (e.g., altitude and speed maintained by Autopilot, forward/backward movement of the flight stick controls the view of the helicopter, the digital map stay oriented automatically, etc.) and then they flew a practice route. The practice run took about 7 to 8 min, giving subjects enough time to get familiar with the simulated environment and the simulator itself.

Following the calibration phase and equipment familiarization navigation route exercise, subjects were briefed on the main navigation route (CleghornWest, Fig. 1) for up to 20 min. After the brief, subjects completed the pre-task questionnaire and then were directed back to the flight simulator and evaluators re-verified calibration. Subjects then flew the main route (6 min long) while evaluators collected eye-scan data and flight information. If a subject went too far off course, the experimenter would verbally intervene, giving them a course to guide the subject back to a waypoint. Subjects then completed the post-task questionnaire and were debriefed. Total experiment time varied from 1 to 1.5 h.

Statistical Analyses

We used Spearman's rank correlation to see if expertise was associated with flight performance and/or visual scan characteristics. We used a significance level of $\alpha = 0.05$ for determining whether to reject the null hypotheses, though here we report all the *P*-values for those who might prefer an alternate significance level. For a regression analysis on the easy route section between TFH and gaze parameters, we refer the reader to Sullivan et al. (15).

The main outcome measures for the flight and navigation performance were 1) RMS error of the flight trajectory; and 2) navigation accuracy, i.e., whether pilots were on-track (within 0.5 km from the waypoints) or off-track (deviated more than 0.5 km). The RMS error is defined as

$$\text{RMS error}_{k,k+1} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^a - x_i^o\right)^2}$$

where for n data points between waypoints k and k+1 x_i^a is the actual flight position and x_i^o is the corresponding reference trajectory point for the i_{th} point.

Navigation accuracy was assessed on the easy route section and the difficult route section, respectively. Navigation accuracy was quantified as a 2 if the pilot was on-track for both legs of the section (e.g., on-track for waypoints 2–3 and waypoints 3–4 in waypoints 2–4), 1 if the pilot was on track for only one leg (e.g., on-track only for waypoints 5–6 in waypoints 5–7), and a 0 if they were off-track for both legs. Being on-track was determined based on whether or not the subject was closely located (threshold was 0.5 km) to designated waypoints and by the subject's debrief. Navigation accuracy is a relaxed variation of the conventional RMS error, which allows acceptable deviation and captures "good-enough" or "satisfying" characteristics of tracking tasks (7,20), whereas RMS error penalizes any deviations from the waypoints.

The main outcome measures for visual scan patterns were 1) median of dwell duration; 2) OTW scan time; 3) number of OTW-Map view changes; and 4) number of fixation points per unit time. Dwell duration (or the duration of fixations) is calculated as a period between consecutive saccades (12). Because the navigation tasks had two different views (OTW and Map), the variables OTW and Map scan time ratio and number of OTW-Map view changes were included to account for how many features pilots scanned per view. Data from faceLAB, X-plane, and the image generator were combined into a text file and all data were processed in MATLAB R2010a. The main outcomes from the survey data were self-reported level of navigation difficulty.

RESULTS

Preliminary Analyses

Spearman's rank correlation is denoted by ρ and the corresponding *P*-value is shown as *P*. As would be expected, TFH was positively correlated with overland flight hours ($\rho = 0.64$ and $P = 0.02$), days since last flight ($\rho = 0.77$ and $P = 0.003$), and days since last overland flight ($\rho = 0.79$ and $P = 0.002$), but not with any other demographic variables, such as age or branch of service. Results from the pre-task surveys indicated that subjects estimated waypoints 5–7 (i.e., the difficult route section as determined by the SME) would be more difficult than waypoints 2–4 (i.e., the easy route section determined by the SME) prior to the navigation task [$t(11) = 3.163$, $P < 0.01$]. After the task completion, subjects still assessed the difficult route as more difficult than the easy route [$t(11) = 8.300$, $P < 0.001$].

Route difficulty affected actual flight and navigation accuracy. As expected, and as can be seen in **Table I**, RMS error increased and navigation accuracy decreased from the easy route section to the difficult route section [$t(11) = 5.171$, $P < 0.001$ and $t(11) = 3.924$, $P < 0.01$, respectively]. For the easy route, 10 pilots were on course, whereas only 3 pilots were on course for the difficult route. These results confirmed the SME's evaluation. Comparing subjects' pre-task estimation with their post-task assessment, we found that pre- and post-reports were consistent for the easy route, but that the difficult route was under-estimated in the pre-task estimate compared to the post-task assessment [$t(11) = 2.901$, $P < 0.001$]. Table I shows mean and SD of each dependent measure on the easy route section and the difficult route section, respectively. Dwell parameters in the helicopter navigation tasks were in the range of results previously reported (17). Also, the distribution of dwell duration was skewed to the left. We, therefore, used the median dwell duration in statistical analyses rather than using mean dwell duration.

TABLE I. MEAN, MEDIAN, AND SD OF DEPENDENT VARIABLES.

	Leg 1 (Easy, Waypoints 2-4)			Leg 2 (Difficult, Waypoints 5-7)		
	Mean	Median	SD	Mean	Median	SD
Navigation accuracy (max = 2.0)	0.92	1.0	0.19	0.62	0.5	0.22
RMS error	11.5 ft	9.05 ft	7.8 ft	30.6 ft	30.5 ft	14.2 ft
Median dwell duration	229.1 ms	215.8 ms	47.3 ms	212.8 ms	208.6 ms	34.1 ms
Median OTW dwell duration	226.5 ms	227.1 ms	38.7 ms	213.9 ms	207.6 ms	43.1 ms
Median Map dwell duration	297.5 ms	224.8 ms	159.0 ms	257.5 ms	230.1 ms	91.4 ms
No. of OTW fixations per view	4.1	3.0	2.6	3.3	2.4	1.8
No. of Map fixations per view	1.74	1.79	0.65	1.78	1.55	0.61
OTW scanning time	61%	60%	12%	56%	56%	9%
No. of OTW-Map view changes per second	1.35	1.34	0.63	1.30	1.20	0.56
Route difficulty estimation (max = 75)	19.4	19.0	7.4	32.5	37.0	11.9
Route difficulty assessment (max = 75)	15.4	13.0	9.9	50.4	47.3	11.1

None of the gaze parameters were significantly different between the two route sections, possibly due to the wide range of variability in all gaze parameters, with the most variability occurring with median Map dwell duration. Of note, the number of fixations per OTW view was more than that of the Map view for both routes [easy route: $t(11) = 3.067$, $P < 0.01$, and difficult route: $t(11) = 3.586$, $P < 0.005$] and OTW scanning time was more than 50% for both routes. This result indicates that, regardless of route difficulty, pilots tend to spend more time looking at and fixating OTW relative to the Map view.

Navigation accuracy was correlated negatively with two gaze parameters on the easy route (median dwell, $\rho = -0.45$, $P < 0.1$; median OTW dwell, $\rho = -0.52$, $P < 0.05$; pilots who were on-track had shorter dwell times on the easy route) whereas no significant correlation with any gaze parameters was found on the difficult route. Spearman's rank correlation coefficients among flight, navigation, gaze, and subjective data were calculated for both route sections and are shown in **Table II**. The lower half of the table corresponds to Spearman's rank correlation coefficient between dependent variables in Leg 1 and the upper half to that of Leg 2. Navigation accuracy was correlated negatively with RMS error, OTW dwell duration, and post-task route assessment on the easy route ($\rho = -0.52$, $P < 0.05$; $\rho = -0.52$, $P < 0.05$; $\rho = -0.55$, $P < 0.05$). On the other hand, navigation accuracy was only correlated negatively with post-task route assessment on the difficult route ($\rho = -0.53$, $P < 0.05$). As would be expected, most gaze parameters were correlated with each other on both the easy and difficult routes; for example, OTW dwell and OTW-Map view changes were correlated negatively in both legs ($\rho = -0.66$, $P < 0.05$ and $\rho = -0.69$, $P < 0.001$, respectively). The subjects' pre-task estimation was correlated negatively with OTW scan duration on the easy route ($\rho = -0.61$, $P < 0.05$) and post-task assessment was correlated positively with number of fixations per Map on the difficult route ($\rho = 57$, $P < 0.05$). Pre-task estimation and post-task assessment were correlated negatively on the difficult route (ρ=-0.69, $P < 0.001$), whereas no correlation was shown for the easy route.

Hypothesis 1

The experimental data did not support either Hypothesis 1a regarding the relationship between TFH and navigation accuracy or 1b regarding the relationship between TFH and RMS error. That is, TFH was not a significant predictor of either navigation accuracy or RMS error for both the easy and difficult route sections. Of course, failure to prove the alternative hypothesis does not mean the null is true, though we note that the lack of association between TFH and RMS error is consistent with our previous work (15) and a post hoc power analysis (see the Discussion section) indicates the sample size provided reasonable power to detect correlations similar to those observed between other factors in this study.

Hypothesis 2

The experimental data supported Hypotheses 2a, c, and e on the association between TFH and gaze parameters. Specifically, TFH predicted median OTW dwell, number of fixations per OTW, and number of OTW-Map view changes on both the easy and difficult route sections ($\rho = -0.66$, $P < 0.01$; $\rho = -0.62$, $P < 0.05$; and $\rho = 0.59$, $P < 0.05$ for easy route sections and $\rho = -0.69$, $P < 0.01$; $\rho = -0.59$, $P < 0.05$; and $\rho = -0.65$, $P < 0.05$ for difficult route sections). **Table III** contains the correlations between TFH and scan parameters by route section. As illustrated in Table III, these results indicate that pilots with more TFH showed a more efficient scan pattern characterized by shorter median OTW dwell, less number of fixations per OTW and more number of

TABLE II. SPEARMAN'S CORRELATION BETWEEN TFH AND GAZE PARAMETERS.

	Easy Route	Difficult Route
Median OTW dwell duration	−0.66**	−0.69**
Median Map dwell duration	−0.47	−0.02
No. of OTW fixations per view	−0.62*	−0.59*
No. of MAP fixations per view	0.12	−0.65*
No. of OTW-Map view changes	0.59*	−0.65*
OTW scanning time	−0.61*	−0.30

* $P < 0.05$, ** $P < 0.01$.

TABLE III. SPEARMAN'S RANK CORRELATION COEFFICIENT ρ BETWEEN NAVIGATION ACCURACY, FLIGHT PERFORMANCE, GAZE PARAMETERS, AND SUBJECTIVE MEASURES.

		Difficult Route										
		Navigation Accuracy	RMS Error	Median Dwell	Median OTW Dwell	Median MAP Dwell	No. of Fixations per OTW	No. of Fixations per Map	Scan Duration OTW	OTW-MAP View Changes	Route Diff Estimation	Route Diff Assessment
Easy Route	Navigation accuracy	-	−0.31	0.03	0.20	−0.20	−0.03	−0.31	0.08	0.08	0.22	−0.53*
	RMS error	−0.52*	-	0.30	0.08	0.22	−0.24	−0.15	−0.01	−0.04	−0.04	0.09
	Median dwell	−0.45	0.03	-	0.82*	0.61**	0.31	0.22	0.62*	−0.69**	0.32	−0.31
	Median OTW dwell	−0.52*	0.13	0.94**	-	0.38	0.34	0.36	0.54*	−0.69**	0.37	−0.27
	Median MAP dwell	−0.39	0.00	0.87*	0.76**	-	0.33	−0.01	0.45	−0.60*	0.22	−0.37
	Num. of fixations per OTW	0.26	−0.27	0.58*	0.45*	0.55	-	0.73**	0.08	−0.78**	0.21	0.06
	Num. of fixations per MAP	0.32	−0.15	0.03	−0.01	−0.07	0.29	-	−0.08	−0.70**	−0.10	0.57*
	Scan Duration OTW	0.13	−0.20	0.43*	0.44	0.44	0.69**	−0.24	-	−0.40	0.23	−0.38
	OTW-MAP view changes	0.00	0.29	−0.78**	−0.66*	−0.73**	−0.83**	−0.29	−0.49	-	−0.07	−0.06
	Route Diff Estimation	0.00	0.10	−0.33	−0.27	−0.10	−0.32	0.25	−0.61*	0.11	-	−0.69**
	Route Diff Assessment	−0.55*	0.33	0.20	0.09	0.38	−0.13	−0.39	−0.08	−0.12	0.20	-

Lower half and upper half of the table corresponds to the easy route and difficult route, respectively.
$^*P < 0.05$, $^{**}P < 0.01$

OTW-Map view changes. Hypotheses 2b, d, and f were not supported, but TFH-by-gaze parameter interactions (e.g., TFH-by-OTW scan duration, TFH-by-fixations per Map view) were found. In particular, TFH was negatively associated with OTW scan duration for the easy route ($\rho = -0.61$, $P < 0.05$), whereas no differences in OTW scan duration were found for the difficult route section. On the other hand, TFH was negatively associated with the number of fixations per Map view only on the difficult route section ($\rho = -0.65$, $P < 0.05$). The interactions suggest that more experienced pilots make subtle changes to their scan pattern when route difficulty increases, where they spend more time scanning out the window and look less often at the map. In contrast, less experienced pilots do not change their scan pattern when navigation difficulty changes.

Hypothesis 3

The hypothesis on an association between TFH and route difficulty estimation was not supported. Regardless of TFH, pilots tended to underestimate the difficult route compared to post-task assessment. Interestingly, pre-task estimation and post-task assessment were negatively correlated for the difficult route ($\rho = -0.69$, $P < 0.001$), which indicates that pilots who estimated the leg to be easy/difficult, after completing the navigation task, then assessed it as more/less difficult, respectively.

As an exploratory analysis, subjects were divided into two groups according to their navigation accuracy (on-track vs. off-track) in both route sections. The purpose of the grouping was to see if on-track subjects can be characterized differently from off-track subjects in terms of gaze parameters. **Table IV** shows dependent measures comparison between these two groups. The descriptive statistics suggest differences between the two groups, but we did not conduct statistical analyses due to the small sample size. Three subjects were in the on-track group and two subjects were in the off-track group. The rest of the subjects had a combination of on- and off-track navigation accuracy, thus they are not included in this exploratory analysis.

DISCUSSION

There are a few possible explanations for the lack of a relationship between TFH and navigation accuracy and gaze parameters. First, there is the possibility of this being an underpowered study due to a small sample size. Post hoc statistical power analysis (3,14) showed that the power of correlation analysis ranges between 0.51 and 0.76 given a sample size = 12, $\alpha = 0.05$, and observed $\rho = 0.52\sim0.69$. Second, TFH may not be an accurate measure of expertise for task specific activities. Even instructor-experienced pilots, an alternative proxy for pilot expertise, did not predict gaze and navigation accuracy on both legs. A better measure of overland navigation expertise may be total overland hours, particularly in this cohort of military pilots, some of whom have most of their flight hours over water. However, overland flight hours did not predict gaze parameters

TABLE IV. MEAN, MEDIAN, AND SD OF DEPENDENT VARIABLES FOR SUBJECTS WHO WERE ON-TRACK OR OFF-TRACK FOR BOTH THE EASY AND DIFFICULT ROUTE SECTIONS.

	On-Track Subjects (Three Subjects)			Off-Track Subjects (Two Subjects)		
	Mean	**Median**	**SD**	**Mean**	**Median**	**SD**
TFH	1780 h.	1600 h	454 h	575 h	575 h	813 h
OFH	867 h	850 h	575 h	50 h	50 h	70 h
RMS error	16.4 ft	14.2 ft	0.5 ft	29.7 ft	29.7 ft	7.0 ft
Median dwell duration	215.9 ms	196.7 ms	36.6 ms	250.9 ms	250.9 ms	8.35 ms
Median OTW dwell duration	228.6 ms	214.6 ms	46.3 ms	245.1 ms	245.1 ms	7.06 ms
Median Map dwell duration	220.3 ms	213.3 ms	52.1 ms	286.2 ms	286.2 ms	15.01 ms
No. of OTW fixations per view	3.7	4.3	1.4	2.2	2.2	0.07
No. of Map fixations per view	1.5	1.6	0.2	1.7	1.7	0.54
OTW scanning time	63%	62%	8.1%	55%	55%	15%
No. of OTW-Map view changes per second	1.3	1.4	0.2	1.25	1.25	0.09
Route difficulty estimation	28.2	29.8	2.7	26.0	26.0	2.1
Route difficulty assessment	27.8	27.3	1.7	42.3	42.3	7.4

Q4 TFH = total flight hours; OFH = ; RMS error = root mean square error; OTW = out the window.

better than TFH either. Third, it could be that the difficult routes were very challenging even for the experienced pilots. Evidence supporting this view is that mean level of navigation accuracy for the difficult route was quite low, 0.62 out of a maximum score of 2.0. Additionally, during the difficult route, more experienced pilots showed a scan pattern that was more representative of a novice scan pattern: longer scan time out the window and fewer fixations. Finally, even the more experienced pilots underestimated how challenging the difficult route would be, suggesting that they were unprepared when confronted with that part of the navigation route.

Other surprising results were that gaze parameters only partially predicted navigation accuracy and changes in route difficulty. Pilots with better navigation accuracy in the easy route had lower median OTW dwell times. As shown in Table I, no significant change was shown in OTW scanning time between easy and difficult route sections. However, increased variability in OTW scanning time during the difficult route could have masked any significant relationship between OTW dwell time and navigation accuracy for this route. These results lead to two questions: 1) can we characterize those pilots who had high levels of navigation accuracy, that is, those that showed task specific expertise; and 2) what types of mistakes were pilots making during the flight?

To address the first question, we compared descriptive statistics between pilots who scored 100% on navigation accuracy across all legs and pilots who had very poor navigation accuracy. Although the sample sizes are too small to reach any general conclusions, the statistics suggest future hypotheses to be tested with larger sample sizes. The high performance pilots are characterized by more THF, more overland hours, lower RMS, shorter overall and map dwell duration, more time spent looking out the window, and more accurate pre-task estimates of route difficulty.

Regarding the second question, whether subjects perceive their whereabouts correctly is critical for successful mission completion. Common frequent visual misperceptions among pilots were observed throughout the study. Some expert pilots successfully located waypoint 6 and made a 90° left turn into a narrow valley toward waypoint 7. However, 9 out of 12 pilots missed this narrow valley mainly due to a field of view angle limitation. Once they passed waypoint 6 without realizing it, another valley appeared on their left. Pilots who missed waypoint 6 made a left turn into this valley, believing they were on track.

As shown in **Fig. 2**, subject 5 missed waypoint 6 and took a left turn into this valley (6′). Then, he flew north

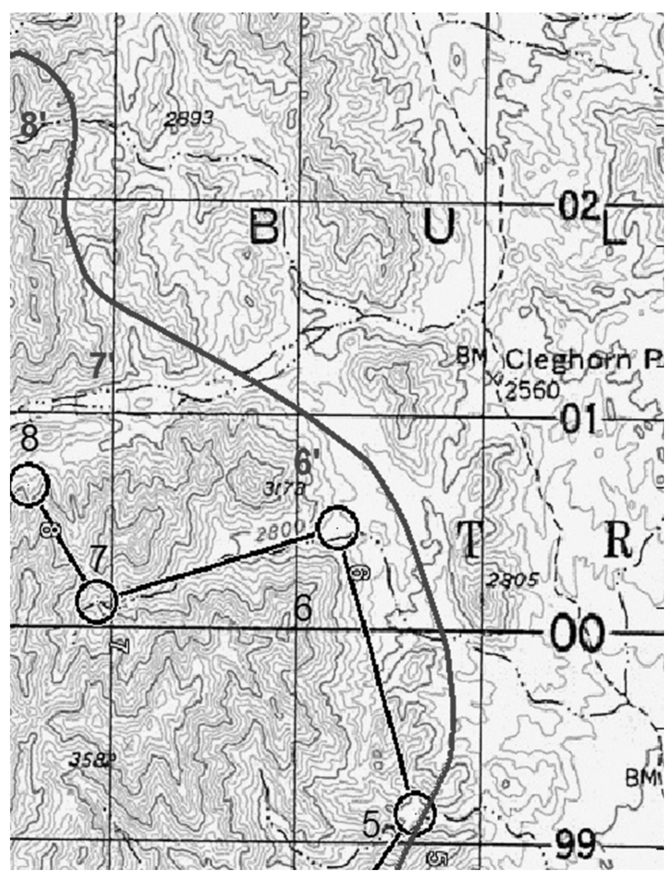

Fig. 2. Subject 5′s actual flight trajectory (continuous line) and planned route (circles with connecting lines) (18).

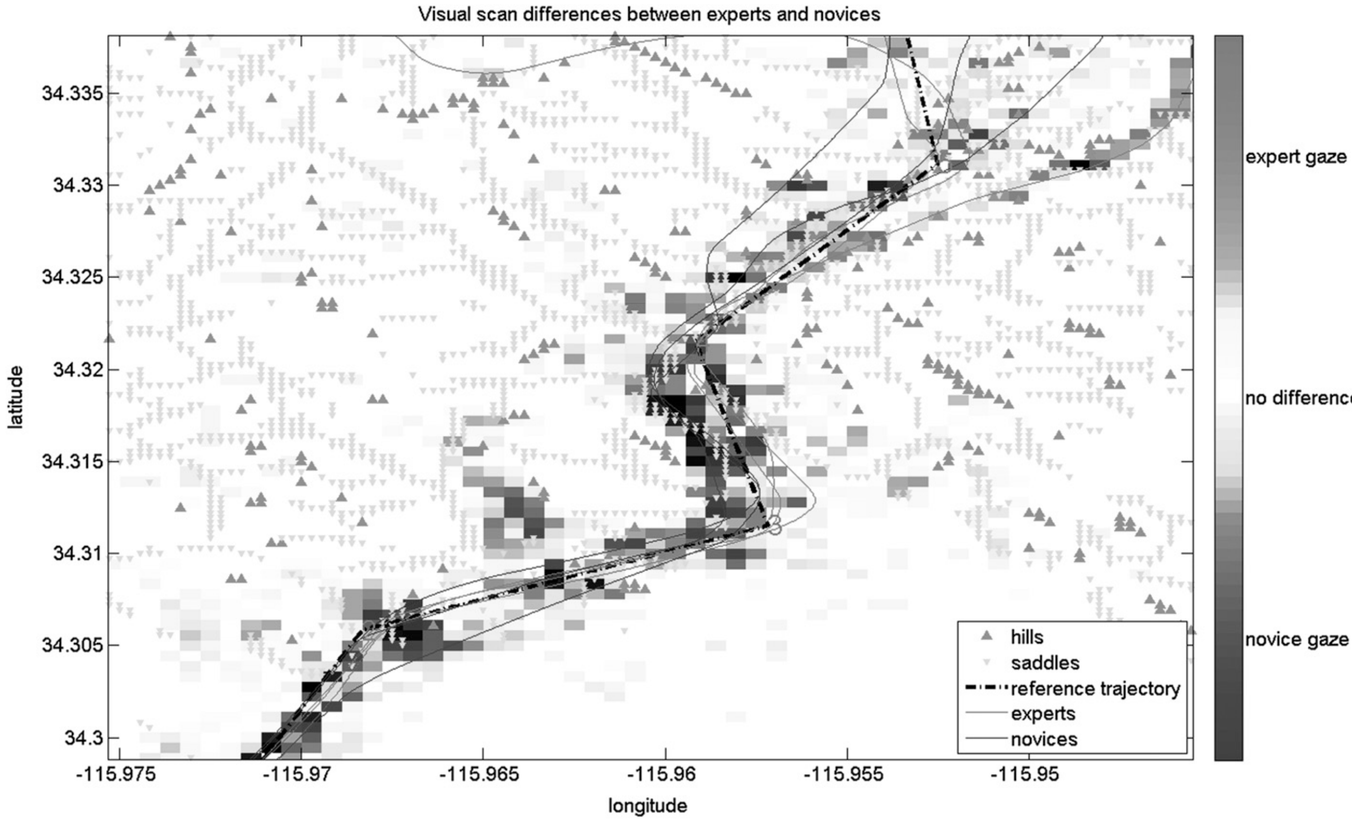

Fig. 3. Visual scan differences between experts and novices. To view this figure in color, please see the online version of this article (DOI: 10.3357/ASEM.3372.2013).

of the intended trajectory (7′ and 8′), believing he was on waypoints 7 and 8. Initially planned waypoints are shown in straight lines connected with circles, whereas the subject's estimation is shown in a thicker, continuous line. On his way from waypoint 6′ to 7′, he saw a valley on the right side of the flight heading in the OTW scene. If he had been on track (i.e., between 6 and 7), he would have been surrounded by hills and should not have been able to see any saddle or valley and his heading would have been much different. Even though his gaze data showed that he scanned the valley, the pilot did not question his orientation. This information indicates the pilot rejected the visual cues that were not compatible with his current belief, which could not have been correct. Thus, the subject did not question his orientation or status, indicating that he overweighed those visual cues that fit into his mental picture by giving little attention (subconsciously) to cues conflicting with it. This type of bias, carrying over initial bias, has also been seen in a cognitive task that tapped inductive biases on cultural evolution (6). Cowden et al. (4) investigated the misperception and showed pilots' perception was wrong 77.86% the time when they were "off-track."

Our gaze pattern analysis thus far has focused on temporal aspects of the data, such as gaze duration, number of fixations, etc. On the other hand, we can also study spatial aspects of gaze parameters, e.g., where in the OTW or Map subjects were looking when navigating. A specific Map scanning strategy used by experts to maintain course was introduced in Sullivan et al. (15). **Fig. 3** shows an OTW gaze histogram of waypoints 2–5 depending on subject expertise sorted by TFH. Red cells indicate the location where experts had more fixations than novices, whereas blue cells represent novice pilots gazing on that area more so than experts. The figure clearly shows that where experts and novices looked were different. For example, experts looked on the left side of the travel direction (hilly terrain) while novices looked on the right (plain area) near waypoint 2. From waypoint 3 to waypoint 4, novices tended to stay and look more to the left while experts tended to look more to the right. The OTW gaze location is, of course, highly subject to helicopter trajectory.

We can conclude TFH predicted gaze parameters, but, in this cohort of military pilots, we cannot reach any firm conclusion regarding the association between TFH and expertise. As future work, how an expert's scan strategy induces better navigation accuracy and how expert pilots obtain the desirable scan strategy should be studied. We should be able to characterize/predict who will perform a task well based on eye gaze pattern vs. those who have scan breakdown. This research is particularly important toward preventing controlled flight into terrain and midair collisions while conducting low-level visual flight rules operations. Scan strategy also differs by task; therefore, a "portfolio" of successful scan strategies by aviation task could be developed.

ACKNOWLEDGMENTS

This work is funded by the Naval Modeling Simulation Office (NMSO) and Office of Naval Research (ONR). We are grateful to Noah Lloyd-Edelman for helping us in calibrating the experimental device and to Michael Day for providing technical support for the experimental setup.

Parts of the results in this paper were presented at the Human Factors and Ergonomics Society 2012 Annual Meeting (19).

Authors and affiliation: Ji Hyun Yang, Ph.D., Quinn Kennedy, Ph.D., Joseph A. Sullivan, Ph.D., and Ronald D. Fricker, Jr., Ph.D., Naval Postgraduate School, Monterey, CA.

REFERENCES

1. Bellenkes AH, Wickens CD, Kramer AF. Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. Aviat Space Environ Med 1997; 68(7):569–79.
2. Bellenkes A, Ford J. Cockpit scan and loss of situational awareness. (Navy aviation training.) Approach June 2002. Available at http://www.highbeam.com/doc/1G1-90189053.html.
3. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Routledge Academic; 1988:109-43.
4. Cowden B, Yang J, Kennedy Q, Sullivan J. Modeling of helicopter pilot misperception during overland navigation. Proceedings of the AIAA Modeling and Simulation Technologies Conference; 13-16 August 2012; Minneapolis, MN. Reston, VA: American Institute of Aeronautics and Astronautics; 2012. Q2
5. Di Nocera F, Camilli M, Terenzi M. A random glance at the flight deck: pilots' scanning strategies and the real-time assessment of mental workload. Journal of Cognitive Engineering and Decision Making 2007; 1:271–85.
6. Griffiths TL, Kalish ML, Lewandowsky S. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. Philos Trans R Soc Lond B Biol Sci 2008; 363:3503–14.
7. Kaneda M, Iizuka H, Ueno H, Hiramatsu M, Taguchi M, Tsukino M. Development of a drowsiness warning system. In: the 14^{th} International Technology Conference Enhanced Safety Vehicles, Paper 94-S3-O-08; 1994; Munich, Germany. Q3
8. Katoh Z. Saccade amplitude as a discriminator of flight types. Aviat Space Environ Med 1997; 68:205–8.
9. Kennedy Q, Taylor JL, Reade G, Yesavage JA. Age and expertise effects in aviation decision making and flight control in a flight simulator. Aviat Space Environ Med 2010; 81:489–97.
10. Klein G. Naturalistic decision making. Hum Factors 2008; 50: 456–60.
11. Marshall SP. Identifying cognitive state from eye metrics. Aviat Space Environ Med 2007; 78(5, Suppl.)B165–75.
12. Morrow DG, Miller LM, Ridolfo HE, Magnor C, Fisher UM, et al. Expertise and age differences in pilot decision making. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2009; 16:33–5.
13. Mourant RR, Rockwell TH. Strategies of visual search by novice and experienced drivers. Hum Factors 1972; 14:325–35.
14. Siegel S, Castellan N Jr. Nonparametric statistics for the behavioral sciences, 2nd ed. McGraw-Hill, 1988.
15. Sullivan J, Yang J, Day M, Kennedy Q. Training simulation for helicopter navigation by characterizing visual scan patterns. Aviat Space Environ Med 2011; 82:871–8.
16. Van Orden KF, Limbert W, Makeig S, Jung TP. Eye activity correlates of workload during a visuospatial memory task. Hum Factors 2001; 43:111–21.
17. Velichkovsky B, Dornhoefer S, Pannasch S, Unema P. Visual fixations and level of attentional processing. In: Proceedings of the 2000 Symposium of Eye Tracking Research &Applications; 2000 November 6-8; Palm Beach Gardens, FL. 79-85. New York: Association for Computing Machinery; 2000.
18. Yang J, Kennedy Q, Sullivan J, Day M. Bayesian modeling of pilot belief and visual misperception in helicopter overland navigation. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics; 2011 October 9-12; Anchorage, AK. New York: IEEE Systems, Man, & Cybernetics Society; 2011.
19. Yang J, Kennedy Q, Sullivan J, Fricker R Jr. Scan patterns on overland navigation in varying route difficulty: is total flight hours (TFH) a good measure of expertise? (In press.) In: Proceedings of Human Factors and Ergonomics Society's 56^{th} Annual Meeting; 2012 October 22-26; Boston, MA. Santa Monica, CA: Human Factors and Ergonomics Society; 2012.
20. Yang J, Mao Z, Tijerina L, Pilutti T, Coughlin J, Feron E. Detection of driver fatigue caused by sleep deprivation. IEEE Trans Syst Man Cybern A Syst Hum 2009; 39:694–705.

Author Query sheet–ASEM3372

Q1 : Author: Is SD correct? Was 'std' short for 'standard deviation'?
Q2 : Author: In Ref. 4, is this the correct information? Was the paper presented at the 2012 AIAA conference?
Q3 : Author: In Ref. 7, please provide the name of the publisher of the conference proceedings and their city and state or country and the year of publication. Please verify that it was the 14th conference and not the 11th.
Q4 : Author: What is OFH short for? Please provide the definition.